УДК 004.93

Л.Г. Ахметшина, А.А. Егоров
Днепропетровский национальный университет им. О. Гончара,
г. Днепропетровск, Украина
Украина, г. Днепропетровск, ул. Научная, 13, akhmlu@mail.ru, egorov@mayak.dp.ua

Влияние вида меры расстояния на чувствительность нейро-фаззи кластеризации многомерных данных

L.G. Akhmetshina, A.A. Yegorov
Dnepropetrovsk National University named after O. Gonchar,
c. Dnepropetrovsk, Ukraine
Ukraine, c. Dnepropetrovsk,Naukova st., 13 akhmlu@mail.ru, egorov@mayak.dp.ua

Influence of the Type of Distance Measure on Sensitivity of Neuro-Fuzzy Clustering for Multidimensional Data

Л.Г. Ахметшина, А.О. Єгоров
Дніпропетровський національний університет ім. О. Гончара,
м. Дніпропетровськ, Україна
Україна, м. Дніпропетровськ, вул. Наукова, 13, akhmlu@mail.ru, egorov@mayak.dp.ua

Вплив виду міри відстані на чутливість нейро-фаззі кластеризації багатовимірних даних

В статье предложен алгоритм гибридной нечеткой кластеризации mdsFCM, который благодаря применению неевклидовых метрик, вычисление которых основано на использовании матрицы ковариации, обладает более высокой чувствительностью при обработке многомерных данных. Представлены экспериментальные результаты применения предложенного алгоритма для кластеризации низкоконтрастных цветных медицинских изображений.
Ключевые слова: многомерные изображения, нейро-фаззи кластеризация, меры расстояний, сегментация, неевклидовые метрики.

This article deals with the description of the hybrid fuzzy clustering algorithm mdsFCM, which is used non-Euclidian distances based on calculation the covariance matrix. This algorithm has the greater level of sensitivity while processing multidimensional data. The experimental results of the application of the proposed algorithm for low-contrast medical color images clustering are shown.
Key words: multidimensional images, neuro-fuzzy clustering, distance measures, segmentation, non-Euclidean metrics.

У статті запропоновано алгоритм гібридної нечіткої кластеризації mdsFCM, який завдяки застосуванню неевклідових метрик, заснованих на використанні матриці коваріації, має більш високий рівень чутливості при обробці багатовимірних даних. Представлені експериментальні результати застосування запропонованого алгоритму для кластеризації низькоконтрастних кольорових медичних зображень.
Ключові слова: багатовимірні зображення, нейро-фаззі кластеризація, міри відстані, сегментація, неевклідові метрики.

Введение

Выполнение кластеризации в том или ином виде часто является неотъемлемой составляющей современных систем обработки и анализа данных. Несмотря на то, что

для решения этой задачи разработано множество методов, они не обладают универсальностью, и их эффективность существенно зависит от специфики исходных данных.

В настоящее время при обработке данных приходится сталкиваться с такими серьезными проблемами:

– большие объемы выборок;

– отсутствие априорных сведений о скрытых в них зависимостях;

– большое количество информативных признаков исследуемых объектов.

Таким образом, используемые алгоритмы кластеризации, с одной стороны, должны быть достаточно эффективны с точки зрения быстродействия, а с другой – уметь обнаруживать неизвестные закономерности без длительной процедуры настройки, а также эффективно использовать множество информативных признаков исследуемых объектов.

Достаточно часто для решения задачи кластеризации применяются нечеткие или нейросетевые алгоритмы. В первом случае зачастую используется алгоритм FCM (Fuzzy c-means) [1] по причине относительной простоты и достаточной для многих задач чувствительности. Среди нейросетевых алгоритмов особый интерес представляют сети, обучаемые без учителя, т.к. они не требуют длительной процедуры настройки, в частности, самоорганизующаяся карта Кохонена (SOM), которая отличается простотой архитектуры и относительно высоким быстродействием [2].

В последнее время популярным является нейро-фаззи подход, который предполагает объединение алгоритма нечеткой кластеризации и нейронной сети в пределах одного метода, что позволяет повысить чувствительность или быстродействие. В качестве примера, реализующего такой подход, можно привести алгоритм гибридной нечеткой кластеризации sFCM [3], [4].

Все вышеперечисленные алгоритмы могут применяться для обработки одномерных и многомерных данных с сопоставимой эффективностью, а следовательно, не используют специфику объектов, имеющих множество информативных признаков.

Существуют алгоритмы нечеткой кластеризации, позволяющие учитывать особенности многомерных данных за счет применения неевклидовых метрик, основанных на вычислении матрицы ковариации – Густафсона-Кесселя и FMLE (Fuzzy Maximum Length Estimates) [5]. Причем применение их для одномерного случая сопоставимо по эффективности с алгоритмом FCM, однако требует больших вычислительных затрат при обработке многомерных данных.

Постановка задачи

Интерес представляет использование в рамках нейро-фаззи подхода нечеткого алгоритма, основанного на применении неевклидовых расстояний.

Таким образом, **целью данной статьи** является повышение чувствительности нечеткой кластеризации многомерных данных за счет применения в алгоритме sFCM матрицы неевклидовых расстояний.

Решение задачи

Предложенный метод гибридной нечеткой кластеризации многомерных данных mdsFCM состоит из одиннадцати шагов.

1. Инициализация числа кластеров c, значения m (экспоненциальный вес нечеткой кластеризации), начальных значений центров нечетких классов v_{fcm}^{0}.

2. Формирование начальных значений весов нейронов SOM, на основании зна-

чений из матрицы центров нечетких кластеров предыдущей итерации v_{fcm}^{t-1}, количество которых выбирается равным $[N_e * c,1]$ (одномерная структура – столбец), где N_e – коэффициент увеличения числа нечетких кластеров c. Каждое новое значение матрицы v_{fcm}^{t-1} вычисляется методом пропорционального распределения [6].

3. Кластеризация исходных данных с помощью SOM, происходящая в два этапа: грубая и тонкая настройка весов нейронов. При этом выполняется динамическое уменьшение количества нейронов, которое осуществляется следующим образом:

3.1. инициализация управляющих параметров нейронной сети:

– вектор радиусов влияния каждого нейрона должен содержать натуральные числа от значения r_1^0 до 1. При этом:

$$r_1^0 = \lfloor N_n^0 / N_e \rfloor, \tag{1}$$

$$r_1^0 = \lfloor N_n^f / max(c, N_e) \rfloor, \tag{2}$$

где N_n^0 и N_n^f – начальные количества нейронов для фаз грубой и тонкой настроек соответственно;

– количество шагов обучения s_{max} равно длине вектора радиусов;

3.2. вычисление параметра, ограничивающего сжатие:

$$p_{\max} = \left(p_{\max}^0\right)^{2-0.5/p_{\max}^0}, \tag{3}$$

$$p_{\max}^0 = \left(N_n^0\right)_{\min} / N_n^0, \tag{4}$$

где $\left(N_n^0\right)_{\min}$ – количество нейронов после предварительного сжатия, выполняемого согласно пунктам 6.1 – 6.5, 6.8, 6.9; $N_c = c / \left(2 * N_n^0\right)$, а минимально допустимое число нейронов $c_{\min} = c$.

3.3. шаг обучения SOM:

3.3.1. попытка уменьшения числа нейронов, если $s / s_{max} < p_{max}$,

где $s \in \{1,...,s_{max}\}$ – номер текущей итерации обучения; $N_c = \dfrac{c * p_{max}}{2 * N_n^0}$,

$$c_{min} = c + \left| \left(N_n^0 - c\right) * \left(1 - s / s_{max}\right) \right|. \tag{5}$$

Применение выражения (5) позволяет ограничить уровень сжатия количества нейронов в зависимости от номера итерации;

3.3.2. при уменьшении числа нейронов вычисляется максимальное значение радиуса влияния текущей итерации r_1^i по формуле

$$r_1^s = \left(2 * r_1^{s-1} * N_n^s\right) / \left(N_n^{s-1} + N_n^{s-2}\right), \tag{6}$$

где r_1^{s-1} – максимальное значение радиуса влияния каждого нейрона на предыдущей итерации; N_n^{s-1} и N_n^{s-2} – количества нейронов на двух предыдущих итерациях

соответственно, причем первоначальные значения этих параметров $N_n^{s-1} = N_n^{s-2} = N_n^0$;

3.3.3. вычисление разности $\Delta r_1 = r_1^{s-1} - r_1^s - 1, (\forall s \in \{2, \ldots s_{max}\})$. Если $\Delta r_1 > 0$, то номер текущей итерации увеличивается на значение Δr_1;

3.3.4. на основе Евклидовых расстояний для каждого экземпляра исходных данных определяется нейрон-победитель;

3.3.5. модификация весов нейронов-победителей и соседних с ними нейронов, определяемых на основе функции соседства;

3.3.6. если $s < s_{max}$, то переход к пункту 3.3.1.

4. Получение новых значений центров нечетких кластеров v_{som}^t путем выбора c значимых центров из матрицы весов нейронов, полученной в результате обучения SOM. Этот выбор является нетривиальной задачей и оказывает влияние на чувствительность метода.

5. Вычисление текущих значений функции принадлежности u^t:

$$u_{k,i}^t = \sum_{L=1}^{c} \left[\frac{D_{i,k}}{D_{i,L}} \right]^{\frac{-2}{m-1}} \begin{pmatrix} \forall k \in \{1, \ldots, c\}, \\ \forall i \in \{1, \ldots, n\} \end{pmatrix}, \tag{7}$$

где n – число экземпляров данных, а D – матрица расстояний между экземплярами исходных данных X и центрами нечетких кластеров, которая вычисляется по следующей формуле:

$$D_{i,k} = \sqrt{\left(X_i - \left(v_{som}^t\right)_k\right)^T * A * \left(X_i - \left(v_{som}^t\right)_k\right)}, (\forall i \in \{1, \ldots, n\}, \forall k \in \{1, \ldots, c\}), \tag{8}$$

причем $A = A_k^t, (\forall k \in \{1, \ldots, c\})$. Следует отметить, что способ формирования матриц A_k^t влияет на чувствительность кластеризации.

6. Динамическое сжатие функции принадлежности на основе расстояний между центрами нечетких кластеров происходит итеративно по следующему алгоритму:

6.1. определение матрицы расстояний в q-мерном пространстве между центрами нечетких кластеров d^0 и значений минимального ненулевого $d_{\min}^0$ и максимального $d_{\max}^0$ расстояний;

6.2. автоматическое определение минимально допустимых расстояний между центрами кластеров – d_a^1 и d_a^2. Расстояние d_a^1 определяется на основе статистических характеристик матрицы d^0 следующим образом:

6.2.1. рассчитывается начальное значение расстояния d_a^1 по формуле:

$$\left(d_a^1\right)_0 = d_c^0 + \left(\sum_{j=1}^{n_d^0} \left(d_c^0 - d_j^0\right) \right) / n_d^0, \tag{9}$$

причем $d_c^0 = min(\overline{d^0}, 0.5 * d_{max}^0)$, где $\overline{d^0}$ – среднее по матрице d^0.

6.2.2. если не существует элементов d_i^0 матрицы d^0, удовлетворяющих условию:

$$d_i^0 < d_c^{L-1} + \left(\sum_{j=1}^{n_d^{L-1}} \left(d_c^{L-1} - d_j^0 \right) \right) / n_d^{L-1}, (L > 0), \tag{10}$$

где $d_c^{L-1} = min\left(\overline{d^{L-1}}, 0.5 * max(d^{L-1}) \right)$, причем $\overline{d^{L-1}}$ – среднее по вектору d^{L-1}, который состоит из n_d^{L-1} элементов d_i^0 матрицы d^0, удовлетворяющих условию (10), то выход;

6.2.3. рассчитывается текущее значение расстояния d_a^1 по формуле:

$$\left(d_a^1 \right)_L = d_c^L + \left(\sum_{i=1}^{n_d^L} \left(d_c^L - d_i^L \right) \right) / n_d^L, (L > 0), \tag{11}$$

где d_c^L вычисляется аналогично значению d_c^{L-1};

6.2.4. если условие:

$$n_d^{L-1} / n_d^L < d_c^{L-1} / d_c^L, \tag{12}$$

выполнилось на предыдущем шаге (кроме 1-го), то выход;

6.2.5. переход к пункту 6.2.2.

Если выход из вышерассмотренного алгоритма произошел из-за невыполнения условия (10), и количество итераций было не менее 2, то расстояние d_a^1 вычисляется по следующей формуле:

$$d_a^1 = max\left(k_{d_c}, 1 - k_{d_c} \right) * \left(d_a^1 \right)_L + min\left(k_{d_c}, 1 - k_{d_c} \right) * \left(d_a^1 \right)_{L-1}, \tag{13}$$

причем $k_{d_c} = d_c^L / d_c^{L-1}$.

Расстояние d_a^2 вычисляется на основе изменения производной [4].

6.3. Выбор минимального $d_{\min}$ из расстояний d_a^1, d_a^2 и d_u, что позволяет избежать негативного влияния от использования слишком большого значения параметра N_c (если $N_c > 0{,}1$, то, как правило, выбирается d_a^1 или d_a^2). Значение d_u вычисляется по формуле:

$$d_u = \overline{\Delta d^0} * N_c + \left(d_{min}^0 \right)^{1 + sqn\left(1 - d_{min}^0 \right) * min\left(d_{min}^0, d_{min}^0 / d_{max}^0 \right)}, \tag{14}$$

где N_c – параметр, задаваемый в ручном режиме и используемый для регулирования степени сжатия, который может изменяться от 0 до 1; $\overline{\Delta d^0}$ – среднее по значениям Δd^0 текущего и первого шагов алгоритма, причем $\Delta d^0 = d_{max}^0 - d_{min}^0$;

6.4. если $d_{\min} = 0$ или не существует расстояния между двумя нечеткими кластерами, меньшего $d_{\min}$, то переход к пункту 6.9;

6.5. изменение каждого центра кластера, отличного от двух выбранных:

$$\left(v^t_{fcm}\right)_k = \left(v^t_{fcm}\right)_k * (1-\Delta_d) + \left(v^t_{fcm}\right)_d * (\Delta_d), \quad (15)$$

$$\Delta_d = \left(d^0\right)_{d,s} / \left(d^0\right)_{d,k}, \left(k \neq d, k \neq s\right), \quad (16)$$

где d, s – индексы выбранных для сжатия удаляемого и модифицируемого центров соответственно; $\left(d^0\right)_{d,s}$ – расстояние между выбранными центрами; $\left(d^0\right)_{d,j}$ – расстояние между удаляемым и одним из оставшихся центров. По формуле (15) модифицируются и центры v^{t-1}_{fcm};

6.6. изменение функции принадлежности $u_{k,i}$ к каждому из кластеров для каждого из векторов исходных данных:

$$u_{k,i} = \max\left(u_{k,i} * (1-\Delta_d), u_{d,i} * (\Delta_d)\right), \quad (17)$$

где $u_{d,i}$ – функция принадлежности к удаляемому кластеру;

6.7. изменение функции принадлежности (путем выполнения нечеткой операции объединения) для двух выбранных центров кластеров, которое отображается для модифицируемого центра, и сжатие путем отбрасывания функций принадлежности удаляемого кластера;

6.8. преобразование центров v^t_{som} и v^{t-1}_{fcm} в соответствии с пунктом 6.7, причем изменение модифицируемого центра кластера происходит с помощью усреднения;

6.9. если количество нечетких кластеров, полученных после сжатия, больше, чем $c_{\min}$ (как правило, равно 2), то переход к пункту 6.1, а иначе – к пункту 6.10;

6.10. масштабирование полученной после сжатия функции принадлежности к нечетким кластерам для каждого экземпляра исходных данных, чтобы удовлетворялось условие $\sum_{k=1}^{c} u_{k,i} = 1$.

7. Формирование матрицы центров нечетких кластеров v^t_{fcm}, которые будут использованы в начале следующей итерации:

$$(v^t_{fcm})_{k,j} = \left(\sum_{i=1}^{n} \left(u_{k,i}\right)^m \cdot X_{i,j}\right) / \sum_{i=1}^{n} \left(u_{k,i}\right)^m. \quad (18)$$

8. Вычисление значения Δ^t_v – среднего по матрице расстояний между центрами нечетких кластеров v^t_{fcm} и v^{t-1}_{fcm}, а также критериев V^t_{xb} и V^t_{fz}, которые являются показателями Ксие-Биени и нечеткости текущей итерации соответственно, следующим образом [5]:

$$V^t_{xb} = \left(\sum_{k=1}^{c}\sum_{i=1}^{n} \left(u^t_{k,i}\right)^m * \sum_{j=1}^{q} \left(X_{i,j} - \left(v^t_{fcm}\right)_{k,j}\right)^2\right) / \left(n * (d_{min})^2\right), \quad (19)$$

$$V^t_{fz} = \left(\sum_{k=1}^{c}\sum_{i=1}^{n} \left(u^t_{k,i}\right)^2\right) / n, \quad (20)$$

где d_{min} – минимальное Евклидово расстояние между центрами нечетких кластеров.

9. Если выполняется условие $C^t_{fz} \ge C^{max}_{fz}$, причем $C^t_{fz} = V^t_{fz} / V^t_{xb}$, а C^{max}_{fz} – максимальный из коэффициентов C^t_{fz}, то запоминаются следующие значения: $\Delta^{max}_v = \Delta^t_v$, $C^{max}_{fz} = C^t_{fz}$, $u^{max} = u^t$ и $v^{max}_{fcm} = v^t_{fcm}$.

10. Если выполняется условие:

$$\Delta^t_v < \varepsilon \text{ или } \left(\left| V^t_{xb} - V^{t-1}_{xb} \right| < \varepsilon \text{ и } \left| V^t_{fz} - V^{t-1}_{fz} \right| < \varepsilon \right), \tag{21}$$

где ε – пороговое значение, а V^t_{xb}, V^{t-1}_{xb} и V^t_{fz}, V^{t-1}_{fz} – показатели Ксие-Биени и нечеткости текущей и предыдущей итераций соответственно, которые вычисляются по следующим формулам, то осуществляется переход к пункту 2.

11. Если выполняется условие:

$$C^t_{fz} < C^{max}_{fz} \text{ и } (\Delta^t_v > \Delta^{max}_v \text{ или } (\Delta^t_v < \Delta^{max}_v \text{ и } p_{\Delta_v} > p_C)), \tag{22}$$

причем

$$p_C = \frac{\left| C^t_{fz} - C^{max}_{fz} \right|}{max(C^t_{fz}, C^{max}_{fz})} * \frac{1}{C^{max}_{fz} - C^{min}_{fz}}, \tag{23}$$

$$p_{\Delta_v} = \frac{\left| \Delta^t_v - \Delta^{max}_v \right|}{max(\Delta^t_v, \Delta^{max}_v)} * \frac{1}{\left(\Delta^{max}_v\right)' - \Delta^{min}_v}, \tag{24}$$

где C^{min}_{fz} и Δ^{min}_v – минимальные значения параметров C^t_{fz} и Δ^t_v соответственно, а $\left(\Delta^{max}_v\right)'$ – максимальное значение критерия Δ^t_v, то происходит возврат к сохраненным значениям матриц нечеткой функции принадлежностей u^{max} и центров нечетких кластеров v^{max}_{fcm}, которые и являются результатом обучения.

Способ формирования матрицы A^t_k перед применением формулы (8) оказывает влияние на чувствительность кластеризации. С учетом обработки многомерных данных A^t_k формируются на основе матрицы ковариации. В этой работе рассматриваются два подхода к вычислению этих матриц:

1. с применением нечеткой матрицы ковариации. При этом каждая из матриц A^t_k вычисляется по формуле:

$$A^t_k = \left(\rho_k * det\left(F^t_k\right)\right)^{1/q} * \left(F^t_k\right)^{-1}, \tag{25}$$

где ρ_k – константа, отражающая знания о данных, подлежащих группированию (если таких знаний до начала кластеризации нет, то $\rho_k = 1, (\forall k \in \{1,...,c\})$), а F^t_k – так

называемая нечеткая матрица ковариации k-й группы – формируется следующим образом:

$$F_k^t = \frac{\sum_{i=1}^{n}\left(\left(u_{som}^t\right)_{i,k}\right)^m \left(X_i - \left(v_{som}^t\right)_k\right)\left(X_i - \left(v_{som}^t\right)_k\right)^T}{\sum_{i=1}^{n}\left(\left(u_{som}^t\right)_{i,k}\right)^m}, \left(\forall k \in \{1,...,c\}\right), \qquad (26)$$

причем u_{som}^t вычисляется по формуле (7), при использовании которой формирование матриц F_k^t, необходимых для получения A_k^t, осуществляется по формуле:

$$F_k^t = \sum_{i=1}^{n}\left(X_i - \left(v_{som}^t\right)_k\right)\left(X_i - \left(v_{som}^t\right)_k\right)^T, \left(\forall k \in \{1,...,c\}\right). \qquad (27)$$

При этом u^0 вычисляется по формуле (8) с использованием матриц v^0 и $A = I$;

2. на основе стандартной матрицы ковариации. При этом $A_k^t = F_k^t$, а матрицы F_k^t вычисляются по формуле (27), но вместо v_{som}^t используется вектор, составленный из средних значений по каждому информативному признаку исходных данных.

Экспериментальные результаты были получены при обработке различных цветных низкоконтрастных изображений, в том числе медицинских, примером которых служат RGB снимки, приведенные на рис. 1 а) и 2 а), представляющие собой результаты дерматоскопии с целью диагностирования меланомы по визуальным признакам, среди которых основными являются наличие бело-синих структур и пятен неправильной формы.

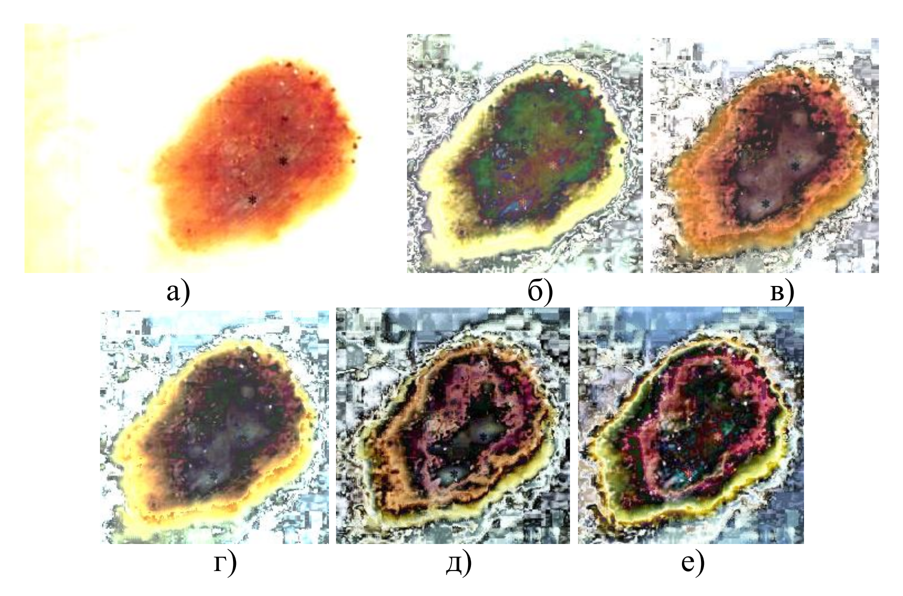


Рисунок 1 – Кластеризация цветного медицинского изображения: а) исходный снимок (330×214 пикселей); фрагмент результатов обработки методами: б) sFCM; mdsFCM с применением: в) нечеткой и г) стандартной матриц ковариации; д) FCM с применением неевклидовых расстояний, $c = 6$; е) Густафсона-Кесселя, $c = 4$

При кластеризации приведенных на рис. 1 а) и 2 а) изображений использовались следующие значения управляющих параметров: $c = 20$ (для алгоритмов sFCM

и mdsFCM); $N_e = 5$ (рекомендуемые значения – 5 или 6, а допустимые значения – целые числа от 2 до 8); после применения карты Кохонена выбирались центры кластеров с максимальным количеством относящихся к ним (имеющих минимальное Евклидово расстояние) векторов исходных данных; $N_c = 0{,}5$; сжатие выполнялось на основе матриц взвешенных Евклидовых расстояний [7] (рис. 1 а) и расстояний Махаланобиса (рис. 2 а). Визуализация результатов нечеткой кластеризации производилась методом сравнения с исходными данными на основе максимального соответствия [8].

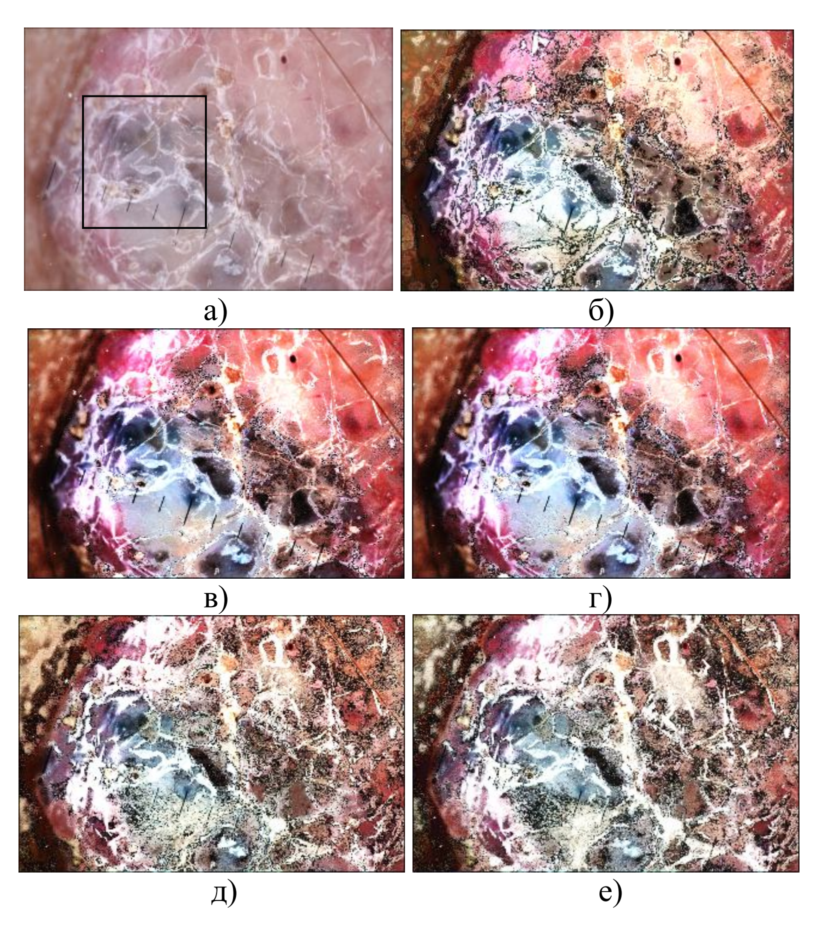
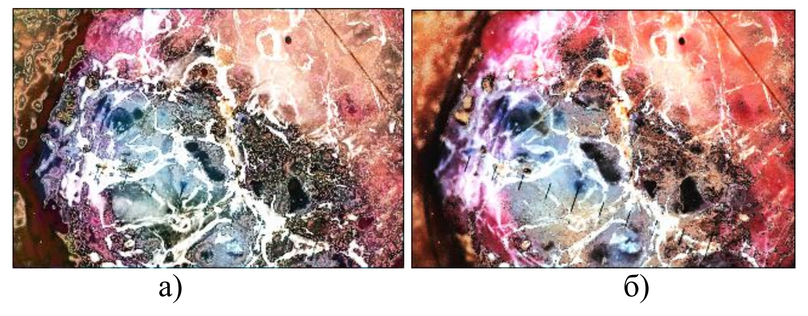


Рисунок 2 – Кластеризация цветного медицинского изображения: а) исходный снимок (323×215 пикселей); результаты обработки методами: б) sFCM; mdsFCM с применением: в) нечеткой и г) стандартной матриц ковариации; д) FCM с применением неевклидовых расстояний, $c = 13$; е) Густафсона-Кесселя, $c = 8$

На изображении, представленном на рис. 1 а), звездочками отмечены области, содержащие очаги меланомы. Применение метода sFCM (рис. 1 б) не приводит к их выделению, в то время как использование алгоритма mdsFCM (рис. 1 в, г) позволяет успешно справиться с этой задачей. При этом очаги меланомы выделяются четче в случае применения стандартной матрицы ковариации (рис. 1 в). На рис. 1 д) и 1 е) приведены результаты применения методов FCM с использованием неевклидовых расстояний, вычисляемых на основе стандартной матрицы ковариации, и алгоритма Густафсона-Кесселя соответственно. При этом количества кластеров $c = 6$ (рис. 1 д) и $c = 4$ (рис. 1 е) соответствуют результатам сжатия для изображений на рис. 1 в) и 1 г). Очевидно, что изображения на рис. 1 д) и 1 е) не обеспечивают четкости выделения очагов меланомы, достаточной для ее однозначного диагностирования.

На изображении, представленном на рис. 2 а), в области, обведенной прямоугольником, содержатся плохо различимые на исходном снимке бело-синие структуры. Применение как метода sFCM (рис. 2 б), так и алгоритма mdsFCM (рис. 2 в, г) позволяет выполнить визуальное диагностирование. При этом метод mdsFCM в обоих случаях позволяет четче выделить цвета в области интереса, что особенно заметно в ее нижней части. За счет улучшения цветопередачи и более четкого выделения структуры снимка появляется возможность повышения точности диагностирования области распространения меланомы. Следует отметить, что из-за черно-белой печати бело-синие структуры в области интереса представляются оттенками серого: черный цвет соответствует темно-синему, а светло-серый – светло-голубому. На рис. 2 д) и 2 е) приведены результаты применения методов FCM с использованием неевклидовых расстояний, вычисление которых основано на стандартной матрице ковариации, и алгоритма Густафсона-Кесселя при $c = 13$ и $c = 8$ соответственно. Очевидно, что изображения на рис. 2 д) и 2 е) не обеспечивают четкого выявления области распространения меланомы.

Уменьшение количества нечетких кластеров для алгоритма sFCM приводит к более существенному снижению чувствительности, чем для mdsFCM. Одним из примеров экспериментального подтверждения этого факта может служить обработка изображения, приведенного на рис. 2 а), этими методами при $c = 4$, результаты которой представлены на рис. 3. Использование алгоритма mdsFCM (рис. 3 б), в отличие от метода sFCM (рис. 3 а), позволяет четче выявить область распространения меланомы. При сравнении с результатами, приведенными на рис. 2 б) и 2 в), изображений на рис. 3 а) и 3 б) соответственно становится очевидным больший уровень снижения чувствительности для алгоритма sFCM, чем для метода mdsFCM.



а) б)

Рисунок 3 – Кластеризация изображения, приведенного на рис. 2 а, при $c = 4$ алгоритмами: а) sFCM; б) mdsFCM с применением стандартной матрицы ковариации

Выводы

Предложенный в данной работе метод гибридной нечеткой кластеризации многомерных данных mdsFCM позволяет повысить чувствительность кластеризации цветных изображений по сравнению с методами sFCM, FCM (при использовании матрицы неевклидовых расстояний) и Густафсона-Кесселя. Выбор способа формирования матрицы ковариации зависит от исходных данных и оказывает влияние на чувствительность кластеризации и время ее выполнения. Кроме того, в ряде случаев удается снизить количество итераций в сравнении с алгоритмом sFCM, что положительно сказывается на быстродействии. Также следует отметить, что при уменьшении количества кластеров чувствительность алгоритма mdsFCM снижается в меньшей степени, чем у метода sFCM.

Литература

1. Леоненков А. Нечеткое моделирование в среде MATLAB и fuzzyTECH / Леоненков А. – СПб. : БХВ-Петербург, 2003. – 719 с.
2. Кохонен Т. Самоорганизующиеся карты / Кохонен Т. ; [пер. 3-го англ. изд. В.Н. Агеева под ред. Ю.В. Тюменцева]. – М. : Бином. Лаборатория знаний, 2008. – 665 с.
3. Ахметшина Л.Г. Сегментация низкоконтрастных изображений алгоритмом гибридной кластеризации SOM-FCM / Л.Г. Ахметшина, А.А. Егоров // Системні технології. – Дніпропетровськ, 2008. – Вип. 2 (55). – С. 34-40.
4. Егоров А.А. Повышение быстродействия метода гибридной нечеткой кластеризации за счет динамического сжатия размерности карты Кохонена // А.А. Егоров / Искусственный интеллект. – 2011. – № 4. – С. 489-497.
5. Рутковский Л. Методы и технологии искусственного интеллекта / Рутковский Л. – М., 2010. – 600 с.
6. Ахметшина Л.Г. Повышение чувствительности гибридной нечеткой кластеризации на основе формирования центроидов пропорционально расстояниям в q-мерном пространстве / Л.Г. Ахметшина, А.А. Егоров // Геометричне та комп'ютерне моделювання. – 2009. – Вип. 24. – С. 193-198.
7. Ахметшина Л.Г. Повышение чувствительности алгоритма гибридной нечеткой кластеризации на основе вычисления взвешенного Евклидова расстояния / Л.Г. Ахметшина, А.А. Егоров // Математичне та програмне забезпечення інтелектуальних систем: IX міжнародна науково-практична конф., (Дніпропетровськ, 16 – 18 листопада 2011 р.). – С. 13-14.
8. Егоров А.А. Визуализация результатов нечеткой кластеризации цветных изображений на основе метода сравнения с исходными данными / А.А. Егоров // Вестник Херсонского национального технического университета. – 2009. – № 2 (35). – С. 195-199.

Literatura

1. Leonenkov A. Nechetkoe modelirovanie v srede MATLAB i fuzzyTECH. S.P.: BHV-Peterburg. 2003. 719 s.
2. Kohonen T. Samoorganizujushhiesja karty M.: Binom. Laboratorija znanij. 2008. 665 s.
3. Ahmetshina L.G. Systemni tehnologii. Dnipropetrovs'k. 2008. Vyp. 2 (55). S. 34-40.
4. Egorov A.A. Iskusstvennyj intellekt. 2011. №4. S. 489-497.
5. Rutkovskij L. Metody i tehnologii iskusstvennogo intellekta. M. 2010. 600 s.
6. Ahmetshina L.G. Geometrychne ta komp'juterne modeljuvannja. 2009. Vyp. 24. S. 193-198.
7. Ahmetshina L.G. Matematychne ta programne zabezpechennja intelektual'nyh system: IX mizhnarodna naukovo-praktychna konf. Dnipropetrovs'k. 1-18 lystopada 2011. S. 13-14.
8. Egorov A.A. Vestnik Hersonskogo nacional'nogo tehnicheskogo universiteta. 2009. № 2(35). S. 195-199.

RESUME

L.G. Akhmetshina, A.A. Yegorov

Influence of the Type of Distance Measure on Sensitivity of Neuro-Fuzzy Clustering for Multidimensional Data

This article deals with the description and experimental verification of the algorithm of the hybrid fuzzy clustering of the multidimensional data, i.e. mdsFCM. This algorithm is the modification of the method sFCM, which correspond the union of the fuzzy clustering algorithm FCM and SOM (Self Organized Map). The essence of the modification consists in the using of the non-Euclidian distances based on calculation of the covariance matrix during fuzzy clustering. This leads to the sensitivity improvement for multidimensional data processing. Two types of the covariance matrix are shown in this article: standard and fuzzy. The experimental test of the proposed clustering method showed the sensitivity improvement and some decreasing of the learning time during multidimensional data processing in comparison with sFCM, FCM (when non-Euclidian distances based on covariance matrix calculation are used) and Gustafson-Kessel algorithms. In the capacity of the initial data, it is used the various low-contrast color medical images, which are the particular case of the multidimensional data. During experimental tests, less sensitivity reducing of mdsFCM algorithm in comparison with sFCM was detected when the number of the fuzzy clusters was decreasing.

Статья поступила в редакцию 01.06.2012.